SPATIAL METHODS IN SCIENCE IMAGE ANALYSIS

Michael Turmon

28 June 2001

A. Scientific Inference

B. Object finding: Volcanoes

C. Image labeling: Sunspots

D. Hierarchical and spatiotemporal models

E. Outlook

Thanks to Michael Burl, Becky Castano, Dennis DeCoste, Eric Mjolsness, and Padhraic Smyth for discussion

turmon@aig.jpl.nasa.gov
http://www-aig.jpl.nasa.gov/home/turmon/

ML FOR SCIENTIFIC INFERENCE

ML methods always give:

Automation: Mechanized process reduces labor and time needed
Cope with increasing data volume (instruments, simulations)
Important for data centers: operations often underfunded
Repeatability: Well-defined algorithm produces results
Uniformity over time key for long-term studies
Allows uniformity among distributed investigators
Crucial for highly charged subjects like climate change

Sometimes one obtains these as well:

Objectivity: Problem-sensitive decision among many conclusions
E.g., model order, number of clusters, which features to use
Often only possible in a limited context or domain
Consensus: Ubiquitous algorithms factor out disagreements
Go beyond ad hoc gadgets to general, cross-domain solutions
Exchange models and algorithms as well as data
Composability: Can analyze machine-generated interpretations
Building a data pipeline, meta-analysis, federated databases

Performance gains are important:

Quality: Quantitative, optimal inference gives better results
Many schemes (implicitly) optimize over interpretations
Gauss obtained the orbit of Ceres by least squares in 1801
Comprehensiveness: Ability to examine more information
Integrate more data within a given interpretation
Achieve total spatial/temporal coverage

GROUND TRUTH — MODEL VALIDITY

Questions brought to fore by scientific problems
 Physical questions that seem decidable in principle...
 ...but whose very intractability motivates inference techniques!

Models for observables

Observables are directly sensed, allowing direct model checks
 Can falsify (Popper 1958), but never fully verify
Computing $P(\text{data} \,|\, \text{model})$ falsifies some models or model classes
 E.g., image modeled as three classes, each of which is normal, is falsified if pooled pixels are not a normal three-mixture

Information on hidden variables

This 'ground truth' is difficult to come by

- Scientists typically cannot identify objects reliably
 Problems become very evident at single-pixel scale
 The most informative test cases are also most uncertain
- Further: Lack of physical understanding of problem means even experts may be surprised at what is really there.

Conceptual inadequacies in models

Methods are often not suitably invariant to resolution
Classes in image segmentation are often not mutually exclusive
Spatial independence is often assumed at some point
Need spatial/temporal stationarity which rarely exists
Bayesian 'dogma of precision': every state can be assigned a probability; every outcome can be assigned a cost (Walley 1991)

SPATIAL MODELS

General References

B. D. Ripley, *Statistical Inference for Spatial Processes*, Cambridge, 1988.
Discrete and continuous random fields; morphological operations

N. A. C. Cressie, *Statistics for Spatial Data*, Wiley, 1993.
Especially strong on geostatistics and models for point-sets

Pattern Theory

U. Grenander and Y. Chow and D. Keenan, *Hands: A Pattern-Theoretic Study of Biological Shapes*, Springer, 1991.
A compelling example of synthesis of a complex shape

U. Grenander and M. I. Miller, "Representations of knowledge in complex systems," *Jour. Roy. Stat. Soc. Ser. B*, 56(4), 549–603, 1994.
Linking abstract models to pixel-level features

Shapes

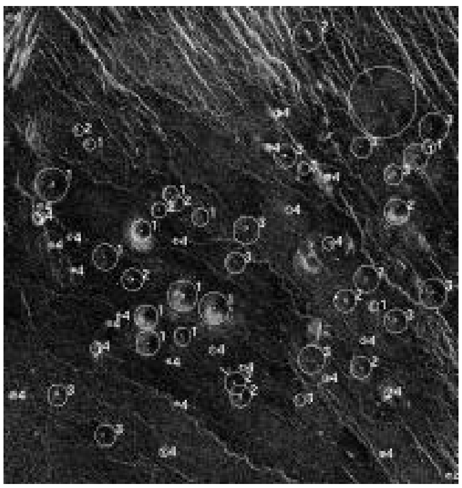
A. Blake and M. Isard, *Active Contours*, Springer, 1999.
Engineering perspective on parameterizing and tracking boundaries

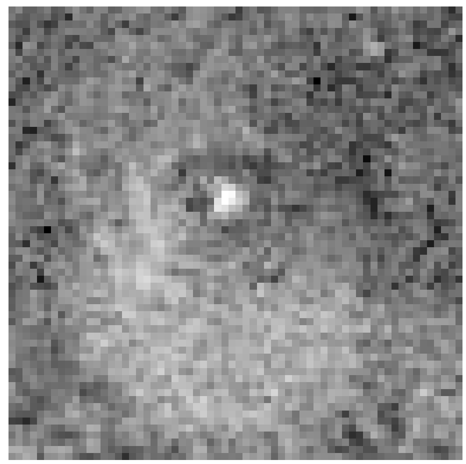
K. V. Mardia and I. L. Dryden, *Statistical Shape Analysis*, Wiley, 1998.
Comprehensive survey of representations and distributions for shapes

OBJECT LOCATION

Known object

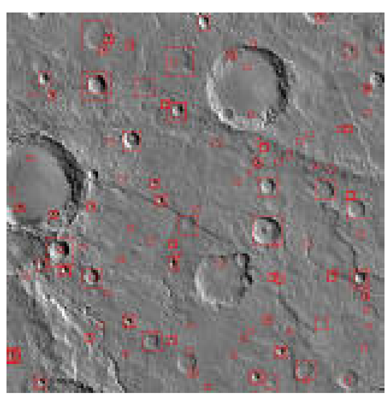
Example: volcanoes on Venus in SAR imagery from Magellan





Known object family

Example: craters (scale variation; also overlap)



Unknown objects

Potential to detect local variations in a background

LEARNING SCHEME

Due to Michael Burl (JPL) and collaborators
(P. Smyth, U. Fayyad, P. Perona)

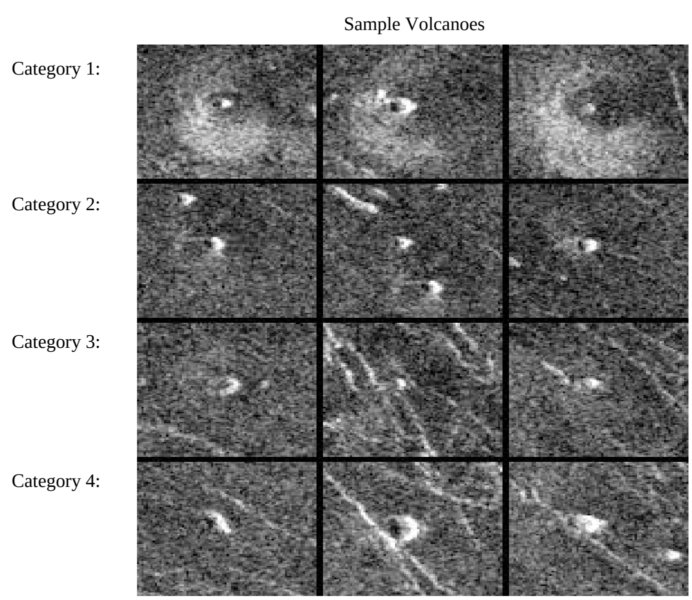
To ease computation, all images reduced in resolution 2×2

Two-phase system
Focus of attention (FOA): Identify all likely candidates
Classification: Assign candidates to classes

FOA sweeps whole image, identifying possible volcano sites which are extracted as square 'chips'

Classification treats chips as i.i.d. inputs and classes as volcano or non-volcano

Training is done with scientist-supplied training chips.



FOCUS OF ATTENTION

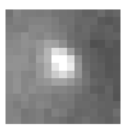
Motivation:

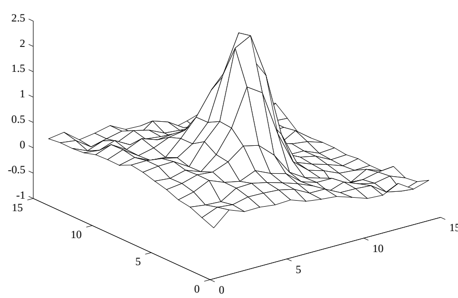
Reduce computation by giving up early on unlikely sites

Allows use of traditional iid classifiers in phase two

Uses scientist-identified volcano sites to find matched filter F

Average of positive examples





F is swept over image to identify strong matches

Less computation by using $F \approx \sum_i f_i f_i^T$

Threshold the correlation to identify potential volcano sites

Sites within four pixels are aggregated in the final list

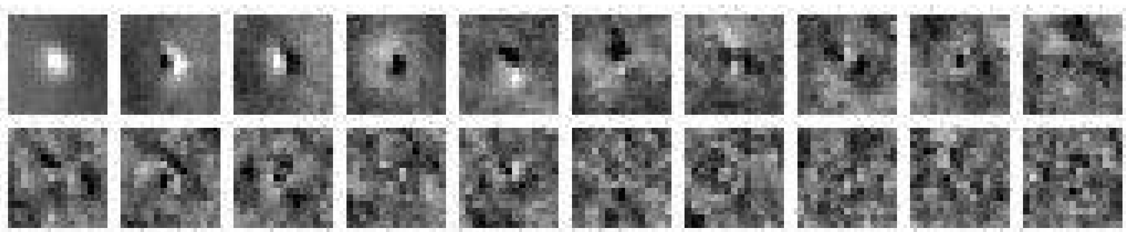
Use family of filters: only limited improvement

CLASSIFICATION

This step is in the realm of classical iid-input algorithms

Non-volcano class is by nature not localized; volcano class is relatively local

Feature selection using PCA compresses 15^2-dimensional data



Quadratic discriminant analysis forms baseline decision rule
- Class-conditional normals with certain mean and covariance
- (μ_v, Σ_v) fitted from volcano training data
- (μ_{nv}, Σ_{nv}) fitted from non-volcano training data
- Classify by thresholding $P(x; \mu_v, \Sigma_v)/P(x; \mu_{nv}, \Sigma_{nv})$

K-Nearest Neighbors a similar-performing alternative
- Use all volcano and non-volcano training chips
- Majority class among the K neighbors of an input chip wins
- Neighbors via weighted Euclidean distance $(x - y)^T R(x - y)$
- R chosen to emphasize pixels close to chip center

Resulting accuracy is about as good as human experts in homogeneous data; degrades markedly in heterogeneous regions

Key seems to be to have good information on local non-volcanoes

DIVIDE AND CONQUER

Schema

Method fits relatively well into Dietterich framework
    Window, decide, merge

FOA algorithm is where all spatial processing happens
    Cleverly, does not choose a fixed window position
    Input scale is the 15×15 pixel window
    Combination rule: FOA-sites within four pels are aggregated
    Output scale is just the granularity of a single volcano

Classification then proceeds independently at each site

Include final V/NV decision into framework as well?
Indicates alternate algorithm where multiple FOA sites are passed through to final classification; then these classes are merged

Fundamental reason this was easy: the discrete, nonoverlapping character of volcanos simplifies the merge

Agenda

Burl et al. 1998: multiple components “make overall system optimization difficult if not impossible given finite training sets”

Optimization seems to enter somewhere, like it or not

IMAGE LABELING

Solar imagery

Reliably identify structures in the photosphere
 Sunspots: Depressed intensity and high magnetic flux
 Faculae: Regions of enhanced intensity and moderate flux
 Quiet sun: everything else
Relate these structures to irradiance changes (weather/climate)
Also: space weather (identify large δ-spots which cause flares)

Mars Geology

Identify soil structure (dust, sand, pebbles)
Detect rocks on soil background
Classify rock types (sedimentary/igneous, weathering, impact)

Methods

Automatic, objective classification using statistical model

Model quantifies the uncertain relation of observables to classes

Model uses spatial information to choose labels

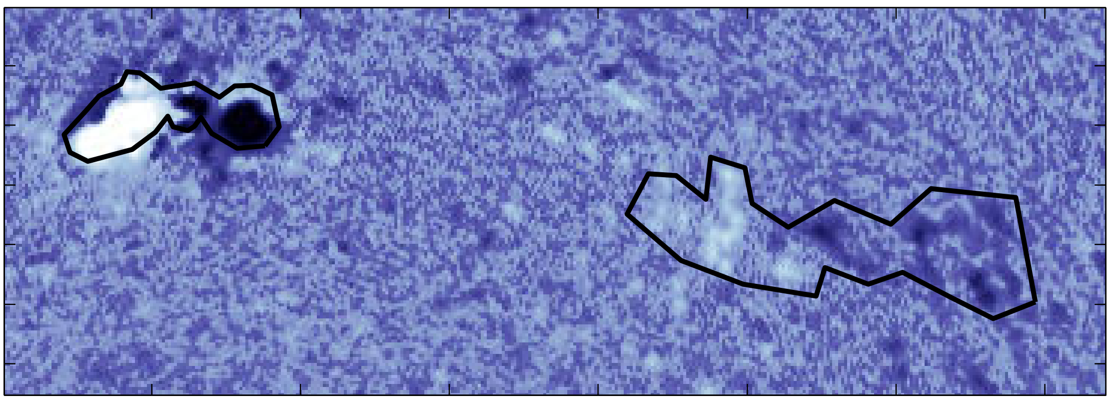
Falsifiable models (Popper 1958) can be checked against
 the data they claim to model

General method that extends unchanged to other settings, e.g.
 more observables
 different number of features
 explicit accounting for miscalibration; outliers
 inclusion of physical knowledge (like sensor noise)

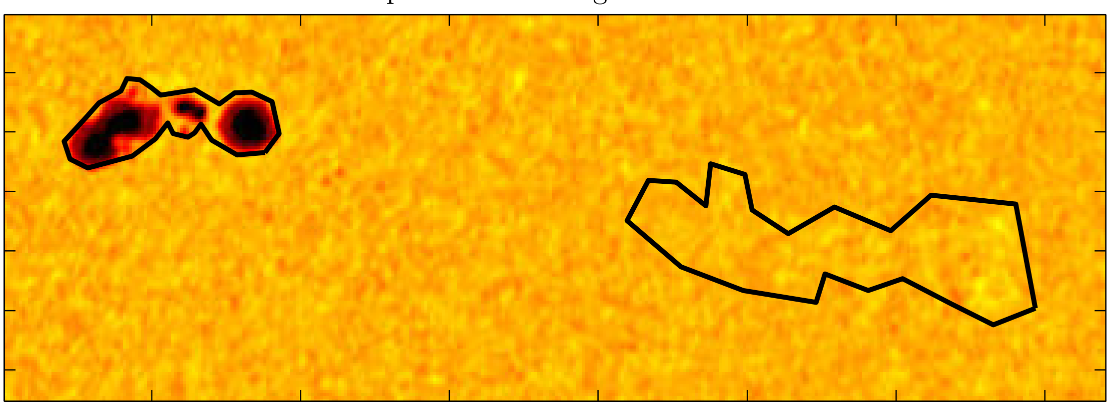
EXAMPLE SOLAR DATA

Irregularly-sampled time series of (full-disk) images
Analyzed May 1996 – Sep 2000; 60 GB across 25 000 images
Below: SoHO/MDI, 17:58 UTC on 7 September 1997

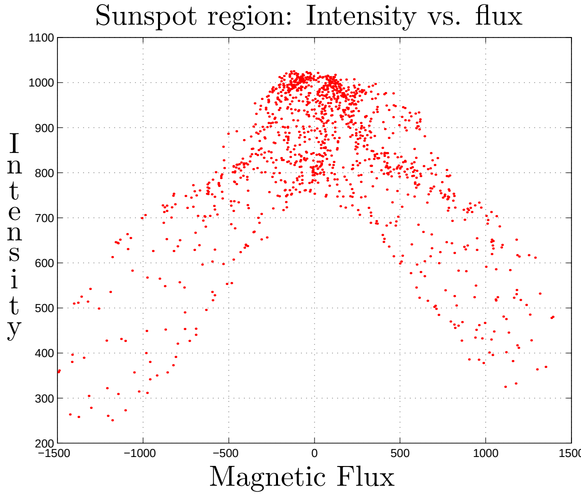
Preprocessed Magnetogram: Detail



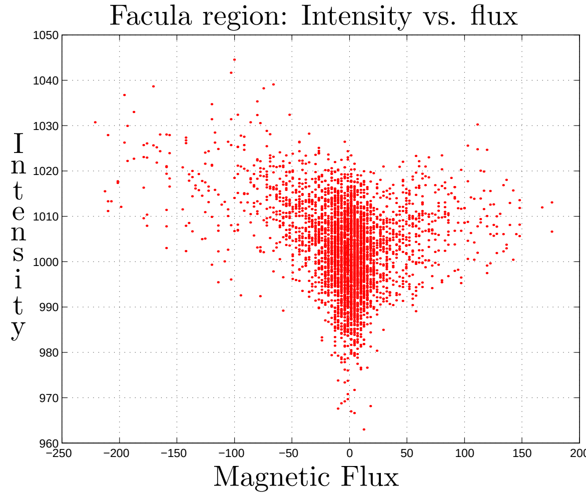
Preprocessed Photogram: Detail



Sunspot region: Intensity vs. flux

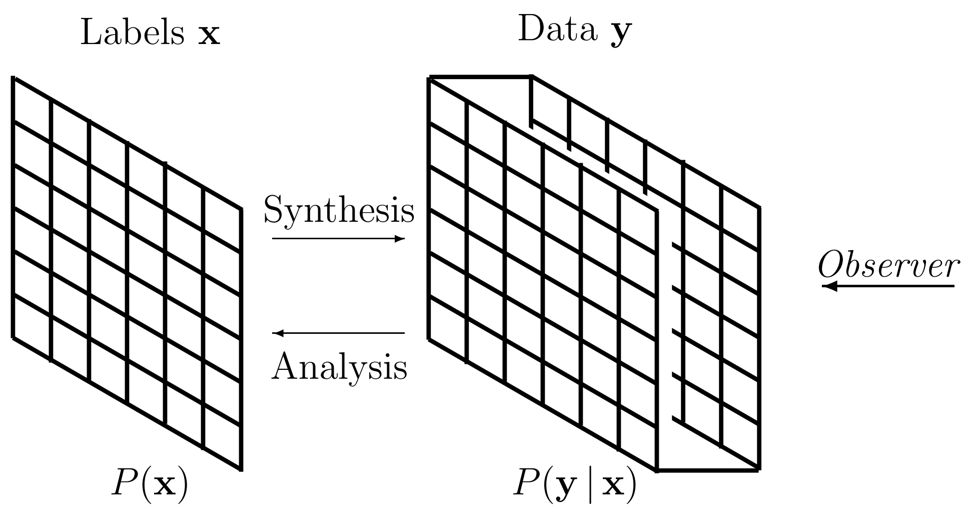


Facula region: Intensity vs. flux



PROBABILISTIC IMAGE MODELS

Quantitatively describe the uncertain relation between observables and labels in a general probabilistic framework



At each spatial position, one of K physical processes is dominant.

Observables arise depending on the dominant physical process.

Generation of observables may be viewed as adding uncertainty (noise) to the underlying dominant process.

Goal of analysis is to invert this noisy mapping.

Variables of the Model

Index set $\mathcal{N}$ of spatial coordinates $s = (i, j)$

Unobservable labels $\mathbf{x} = [x_s]_{s \in \mathcal{N}}$ & observables $\mathbf{y} = [\vec{y}_s]_{s \in \mathcal{N}}$

x_s: small integer $1 \ldots K$ (e.g., ACR/Fac/QS)

$\vec{y}_s$: real vector (e.g., the pair (magnetic field, light intensity))

Statistical model given by two distributions $P(\mathbf{x})$ and $P(\mathbf{y} \mid \mathbf{x})$

MODEL SPECIFICS: I

Describe the two distributions $P(\mathbf{x})$ and $P(\mathbf{y} \,|\, \mathbf{x})$

Linking to Observables with $P(\mathbf{y} \,|\, \mathbf{x})$

Make the link via scientist-labeled images and distribution-fitting

Alternatively, can infer automatically from data via clustering

Obtain K distributions, one for each feature class

As strawman, put forward per-class normal distributions

$$P(\vec{y}_s \,|\, x_s = k) \sim \mathrm{Normal}(\vec{\mu}_k, \Sigma_k)$$

with $d \times 1$ class means and $d \times d$ covariance matrices.

(QS class, $k = 1$: fits the SoHO/MDI data reasonably well using $\vec{\mu}_1 = [\,0 \;\; 1\,]$ and $\Sigma_1 = (0.01)^2 I$.)

For MDI, the normal distribution is inadequate for all classes:
strongly multimodal
cannot even transform to normality (e.g., with |flux|)
quiet class,e.g., contains superpositions of effects
(supergranulation is discernable in scatter plots)
$\Longrightarrow$ it fails standard statistical tests.
...normal model is thus *falsified.*

We must introduce more realistic data models $P(\vec{y} \,|\, x)$

MODEL SPECIFICS: II

Quantifying Spatial Smoothness with $P(\mathbf{x})$

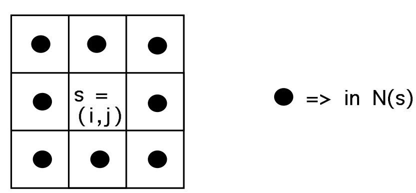
Typically $\beta \geq 0$ controls smoothness in the prior

$$P(\mathbf{x}) = \frac{1}{Z} \exp\Big(-\beta \sum_{s \sim s'} 1(x_s \neq x_{s'})\Big)$$

where $s \sim s'$ means: site s close to site s', e.g. one pixel away

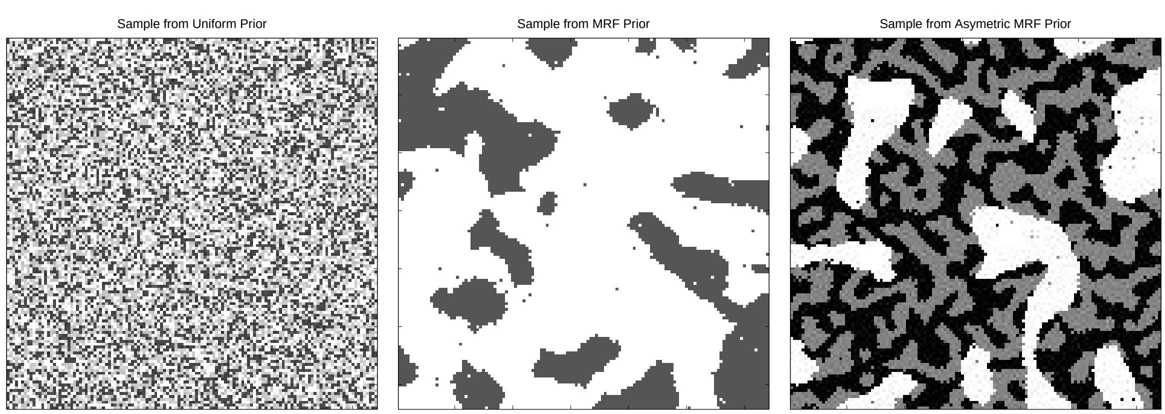
Penalty of β per disagreement of nearby pixels to enforce spatial coherence of labelings

Key property of locality: $P(x_s = x \,|\, x_{(s)}) = P(x_s = x \,|\, x_{\mathcal{N}(s)})$



At $\beta = 0$, penalty and spatial constraint vanish

Sample realizations from $P(\mathbf{x})$



ASIDE: CONTINUITY AND EDGES

Such Markov random field models allow edges in modeled images
Change in discrete hidden variable forces
significant change in real-valued observable

Jumps undesirable in typical image *restoration* contexts

Motivates conditional autoregressive (CAR) model

$$P(x_s = x \mid x_{(s)}) = P(x_s = x \mid x_{\mathcal{N}(s)}) = N(Ax_{\mathcal{N}(s)}, \Sigma)$$

but with conditionally normal distribution

(Autoregression: predict x_s in terms of “itself” $x_{\mathcal{N}(s)}$)

Joint distribution of CAR model is normal, easing computation

Natural parallel with familiar one-dimensional models

	Continuous	Discrete
Time Series	Autoregressive (AR) or Kalman models	Hidden Markov models (HMM)
Imagery	CAR models	Markov random fields (MRF)

MRF computations are the hardest: our best tools do not apply
Non-gaussian, so no reduction to clever matrix manipulations
Bayes net of many short cycles, junction tree algs liable to fail

But: sampling, Metropolis-Hastings, and MCMC methods developed for MRFs enable very complex models

SIMULATING MRFS

Distribution $P(\mathbf{x}) = Z^{-1} \exp\Big(-\beta \sum_{s\sim s'} 1(x_s \neq x_{s'})\Big)$

No direct simulation: no Z, and state space of $\mathbf{x}$ huge!

Randomized algorithm: Gibbs sampler

Basis: craft a MC having P as its stationary distribution
Adaptation of stat-mech methods (c.f. Metropolis *et al.* 1953)
for simulating the state of interacting systems

Iterative algorithm: starts at some labeling and
refines it pixel-by-pixel over many image sweeps

Method:
Choose an initial $\hat{\mathbf{x}}$
Scan pels in raster fashion.
At pel s, find $P(\hat{x}_s = x \,|\, \hat{x}_{(s)})$, $1 \leq x \leq K$. [*]
Choose new $\hat{x}_s$ by drawing from this distribution
Repeat scanning

Result: As scans go to infinity, $\hat{\mathbf{x}} \Rightarrow P(\cdot)$.
That is, iterate enough and the labeling is a draw from $P(\mathbf{x})$

Remarks

Note local combination rule [*]

Flip of one label can eventually influence all labels

This method, and similar Metropolis-Hastings methods, are the
basis for updating more complex spatial models

INFERRING THE LABELING

Invert the noisy data via *maximum a posteriori* (MAP) rule

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y})$$

Bayes formula shows $P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$

For normal $P(y_s \,|\, x_s)$, algebra reveals the objective function

$$\log P(\mathbf{x}|\mathbf{y}) = -\frac{1}{2\sigma^2}\sum_{s\in\mathcal{N}} \left\|\vec{y}_s - \vec{\mu}_{x_s}\right\|^2 - \beta \sum_{s\sim s'} 1(x_s \neq x_{s'})$$

Interpretation

First term: fidelity to data (observation close to its mean)

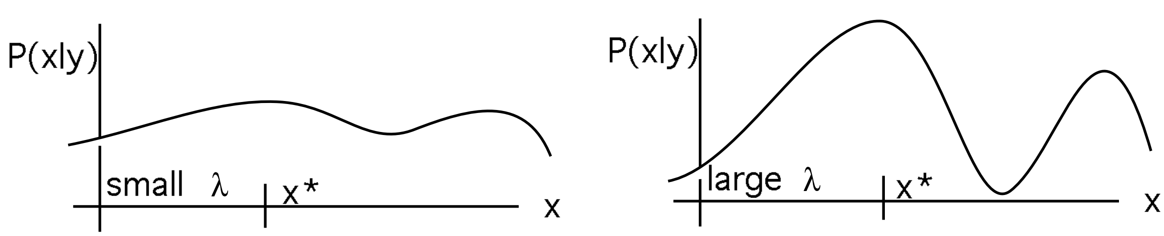
Second term: image smoothness (this couples the pixel labels)

Maximizing $P(\mathbf{x}\,|\,\mathbf{y})$

- Use Gibbs sampler to *draw* from the distribution $P(\mathbf{x}\,|\,\mathbf{y})$
- To *maximize* $P(\mathbf{x}\,|\,\mathbf{y})$, nest G.S. within simulated annealing
  That is, pick large λ and draw via G.S. from

  $$P_\lambda(\mathbf{x}\,|\,\mathbf{y}) := (1/Z_\lambda)\, P(\mathbf{x}\,|\,\mathbf{y})^\lambda$$

  (Effectively scale entire log-posterior, above, by λ)
- Simulated annealing: raise λ as Gibbs sampler iterates
  If λ up slowly enough, mode is reached



- Takes about 3 min/image on Sun workstation (360MHz).

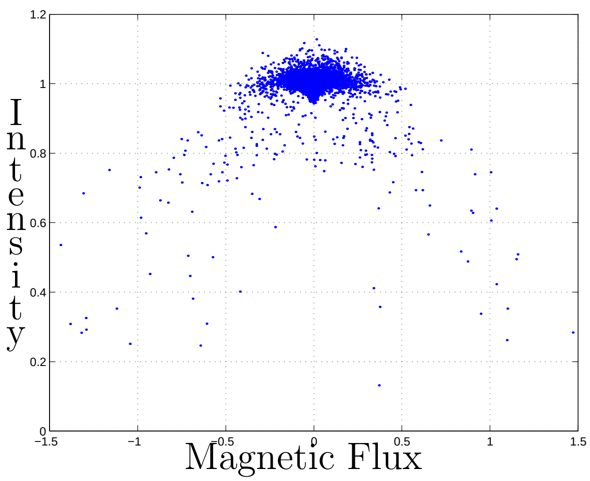
MODELING THE OBSERVABLES

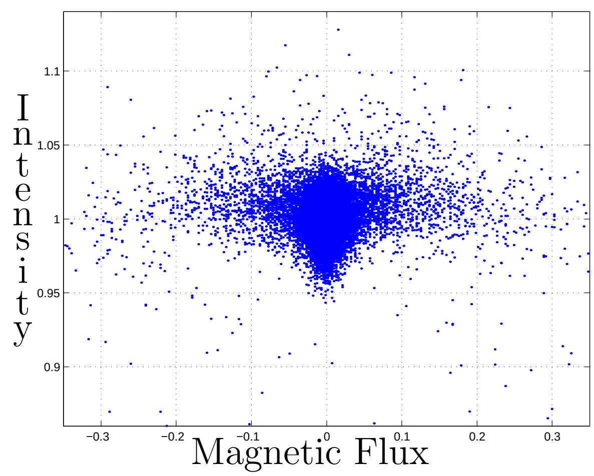
For realistic models, benefit from the flexible mixture density

$$p(\vec{y};\theta) = \sum_{g=1}^{G} \alpha_g N(\vec{y};\, \vec{\mu}_g, \Sigma_g)$$

$$\theta = \{(\alpha_1,\, \vec{\mu}_1,\, \Sigma_1)\cdots(\alpha_G,\, \vec{\mu}_G,\, \Sigma_G)\}$$

Accounts for multimodality and superpositions of effects
A very general family: take G large.





Ask scientists to find regions of type $x_s = k$; estimate θ_k for each

Goal: From data $Y = [\vec{y}^1 \cdots \vec{y}^n]$, find a density model $p(\vec{y};\hat{\theta})$
Method: Determine parameters by maximum-likelihood using Y:

$$\hat{\theta} = \arg\max_{\theta} \log P(Y;\, \theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log p(\vec{y}^i;\, \theta)$$

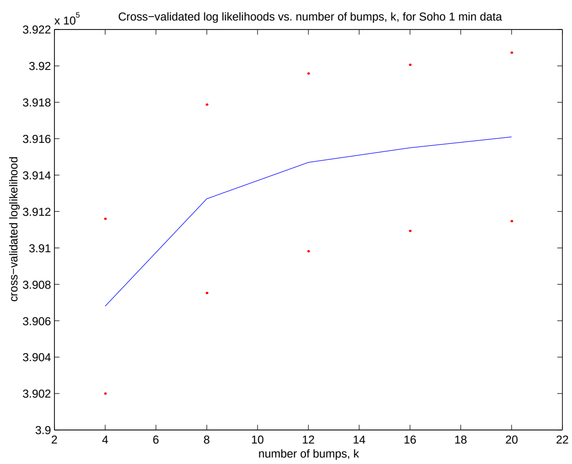
Performed via EM algorithm: done once and the model is fixed

Unsupervised mode: Provide cumulative data over classes, and EM clusters $\vec{y}$ into classes: clusters are extracted after the fact.

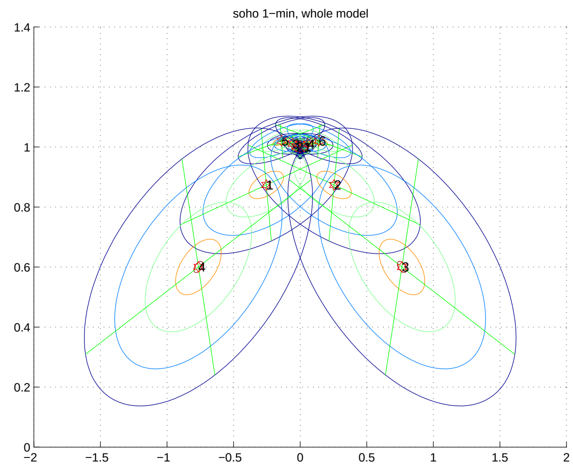
Order selection by cross-validated likelihood (Smyth 1999)

MODELS USED: SOHO/MDI

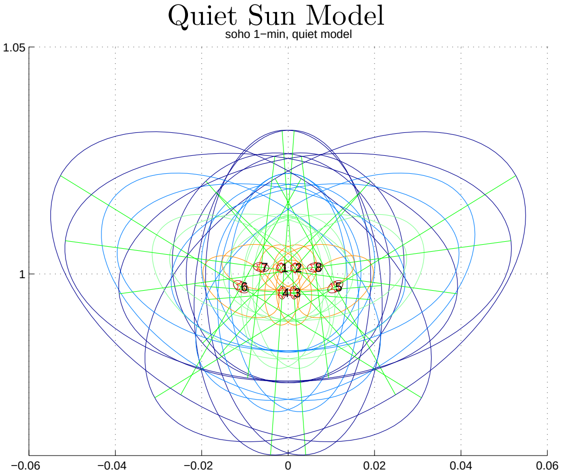
Model Fit, varying Complexity



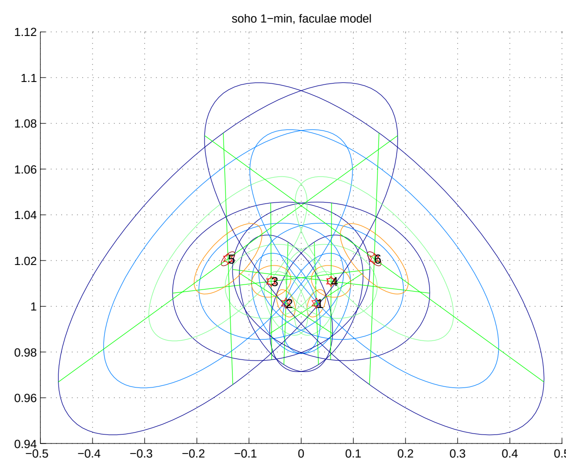
Entire Model



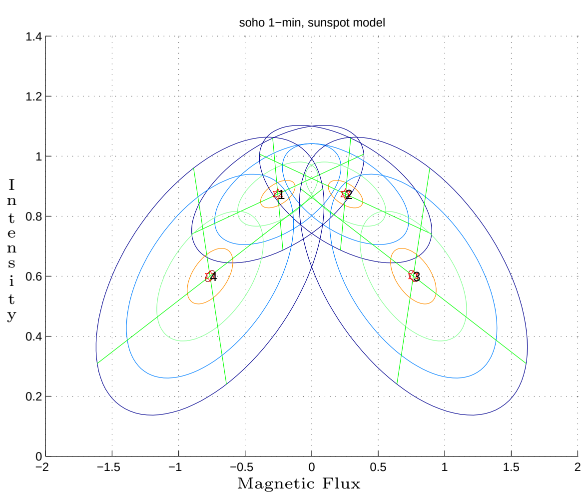
Quiet Sun Model



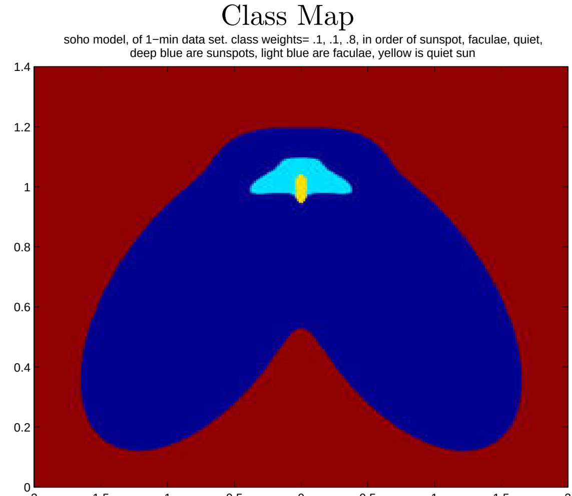
Facula Model



Spot Model

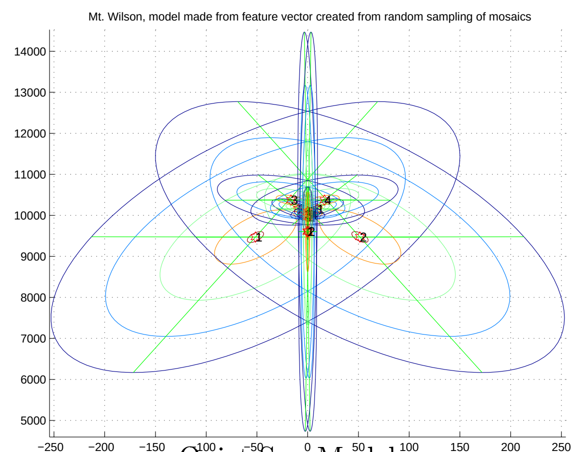


Class Map

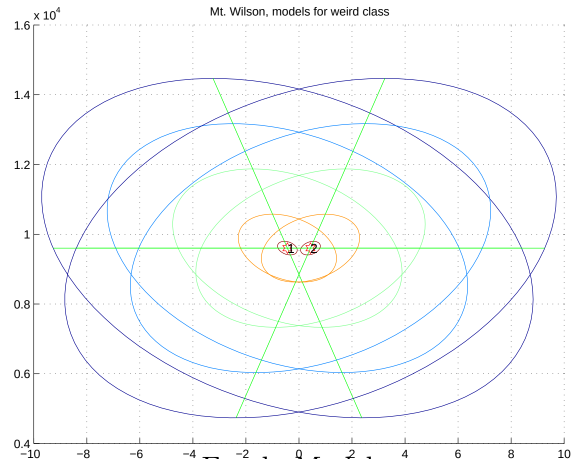


MODELS USED: MT. WILSON

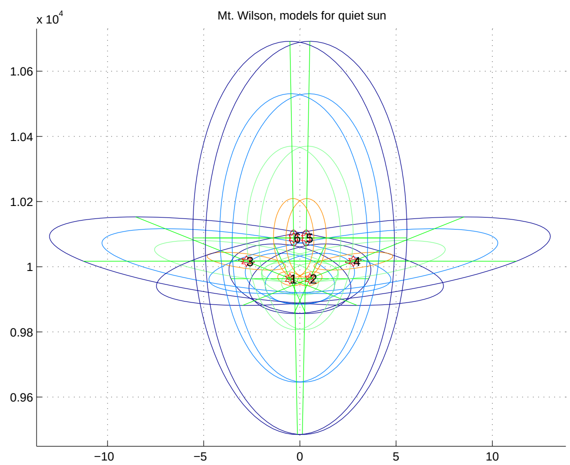
Entire Model



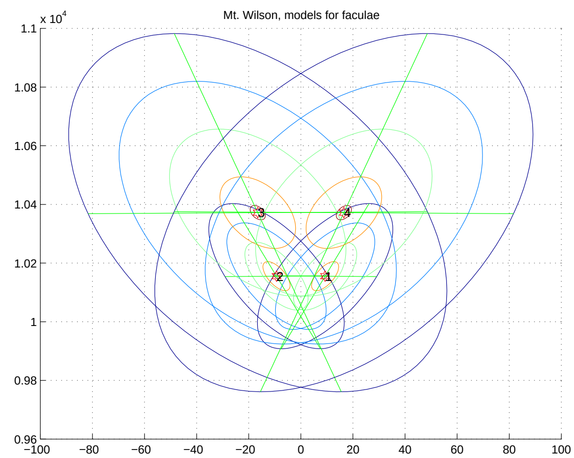
Miscalibration Model



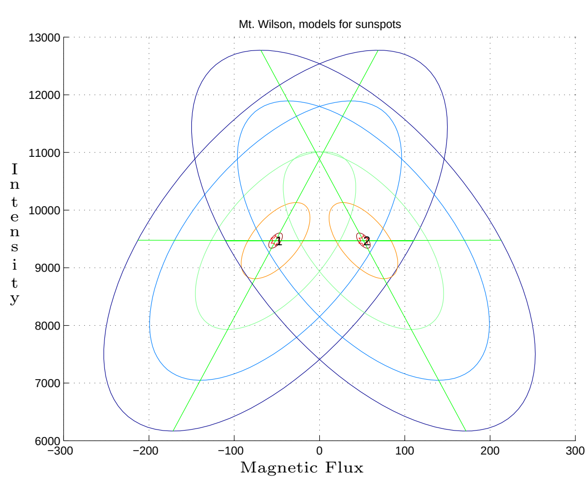
Quiet Sun Model



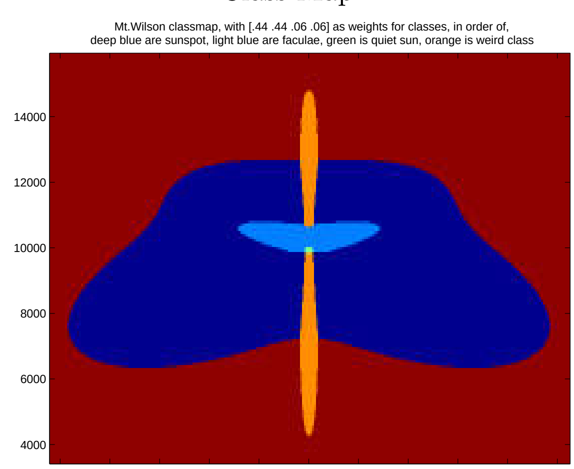
Facula Model



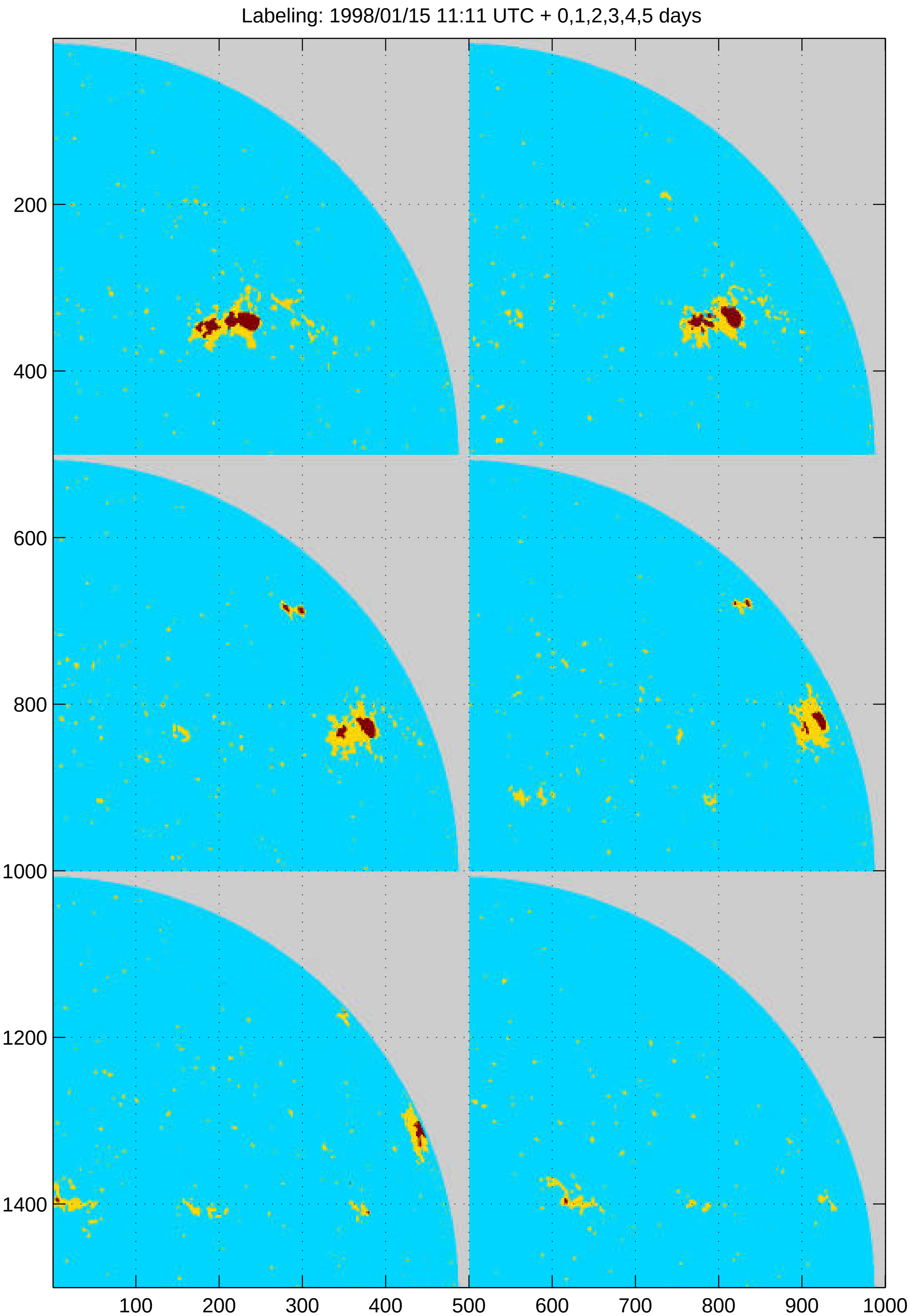
Spot Model



Class Map



LABELINGS



Labeling: 1998/01/15 11:11 UTC + 0,1,2,3,4,5 days
200
400
600
800
1000
1200
1400
100
200
300
400
500
600
700
800
900
1000

HIERARCHICAL SPATIAL MODELS

Better Representations

Represent an object via a *compactly-described* membership function h_s indicating subjective belief site s is active region
— Larger-scale representation of an object
— Provides interpretability

Several Simple Mechanisms

Outlines: Grenander et al., 1991
Polygons: Green 1996
Continuum triangulations: Nicholls 1997, 1998
Delaunay triangulations: Turmon 1998

Binds nearby on-object regions into one object

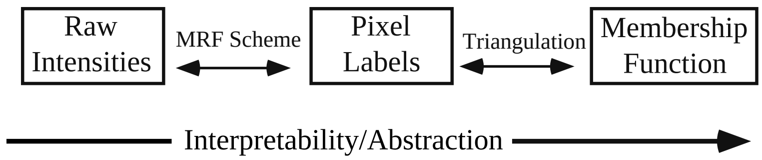
Two fundamental quantities:

Indicator function h_s, $s \in \mathcal{N}$
$h(s) = 1$ means on-object, $h(s) = 0$ if not
Parameterized by tie points in $\mathcal{N}$.
Function complexity $\kappa(h) \geq 0$
e.g., the number of tie points, or
intensity of point process generating tie points

LINK TO OBSERVATIONS

Establish Markov dependence between hierarchical model layers

$$P(h, \mathbf{x}, \mathbf{y}) = P(h)P(\mathbf{x} \,|\, h)P(\mathbf{y} \,|\, \mathbf{x})$$



Probabilistic Model

Penalize complexity by setting

$$P(h) = Z^{-1} \exp\left[-\gamma\, \kappa(h)\right]$$

This choice gives an additive penalty to disjoint objects

Intermediate layer uses h_s to bias the event $\{x_s = \texttt{Object}\}$:

$$-\log P(\mathbf{x} \,|\, h) = \beta \sum_{s \sim s'} 1(x_s \neq x_{s'}) + \alpha \sum_{s \in \mathcal{N}} |1(x_s = \texttt{Object}) - h(s)|$$

The data distribution $P(\mathbf{y} \,|\, \mathbf{x})$ is as before.

- One can do inference by maximizing the posterior

$$P(h, \mathbf{x} \,|\, \mathbf{y}) = P(h, \mathbf{x}, \mathbf{y})/P(y) \propto P(h, \mathbf{x}, \mathbf{y})$$

or minimizing its negative logarithm

$$\gamma\, \kappa(h) + \alpha \sum_{s \in \mathcal{N}} |1(x_s = \texttt{Object}) - h(s)| + \beta \sum_{s \sim s'} 1(x_s \neq x_{s'}) + \frac{1}{2\sigma^2} \sum_{s \in \mathcal{N}} (y_s - \mu_{x_s})^2$$

INFERRING COMPLEX MODELS

We describe inferring shape models for fixed labeling

To speed convergence, replace $1(x_s = \texttt{Object})$ above
with its probability given the data
(Fully analogous to ICE algorithm of Art Owen)

Now the objective simplifies to

$$\gamma\,\kappa(h) + \alpha \sum_{s\in\mathcal{N}} \Big| P(x_s = \texttt{Object}\,|\,\mathbf{y}) - h(s) \Big|$$

Metropolis-Hastings sampler

Inference means choosing tie-point positions

Construct a Markov chain on the state space of tie points

$$\mathcal{V} = \bigcup_k \mathcal{V}_k = \bigcup_k \,(\mathcal{N}\times\mathcal{N})^k$$

that has limit distribution

$$\pi(h) = P(h\,|\,\mathbf{x},\mathbf{y})$$

(Maximize $P(h\,|\,\mathbf{x},\mathbf{y})$ with same annealing setup as earlier)

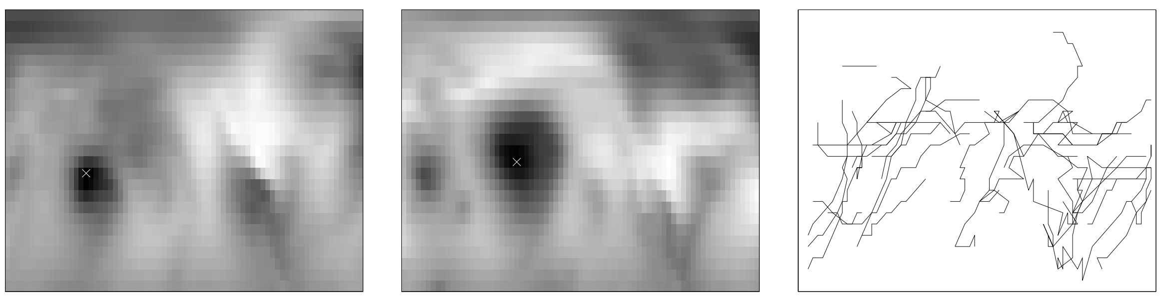
Metropolis-Hastings proposes state changes and probabilistically accepts them to achieve the desired limit distribution

The operator set consists of
  tie-point move (M),
  tie-point raise/lower (R),
  tie-point add (A_k) or kill (A'_k)

SPATIO-TEMPORAL INFERENCE

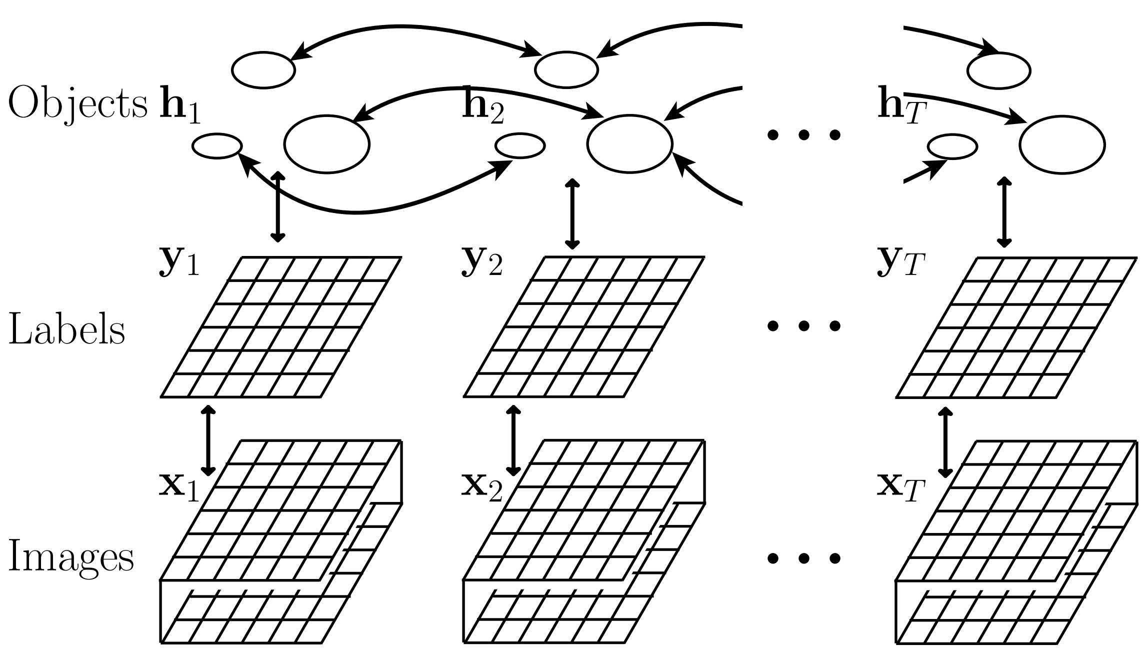
Object trajectories

Sea-level pressure over the Pacific ($\delta t = 48$ hrs.)
Cyclone center shown by white cross
Right: trajectories from a series of (quantized) observations
Data from P. Smyth, UC Irvine



Other examples: sunspot motion, microblock motion from GPS

Objects through time

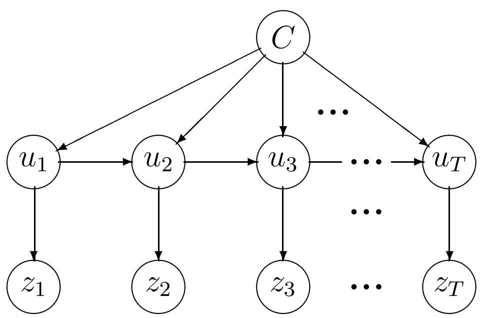


MODELING THE TEMPORAL PART

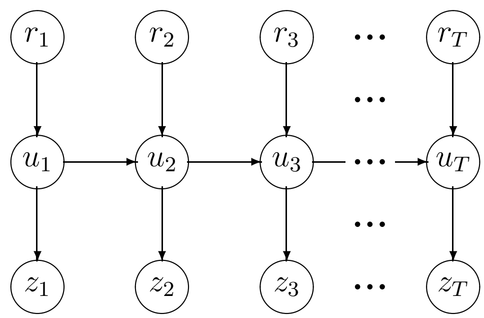
State-based motion models
Include influence of exogenous inputs and observable covariates
Discover motion clusters by uncovering a hidden class C

Examples

Generalizations of the Kalman filter as Bayes nets with state u_t



mixed dynamical model



model with exogenous inputs r_t

Build temporal models atop de-coupled spatial models

Implications

Two domains of divide and conquer
Easy cases: dominant locality in space (sunspots) or time (GPS)
...allows decoupled solutions
Coping with both simultaneously is harder, even beyond current limits of practical optimization technology

Problems...
estimate model parameters automatically
learn the model structure automatically

SPATIO-TEMPORAL MODELING (I)

Base concept of random vector is inadequate
Capture concept of variables on structured index sets

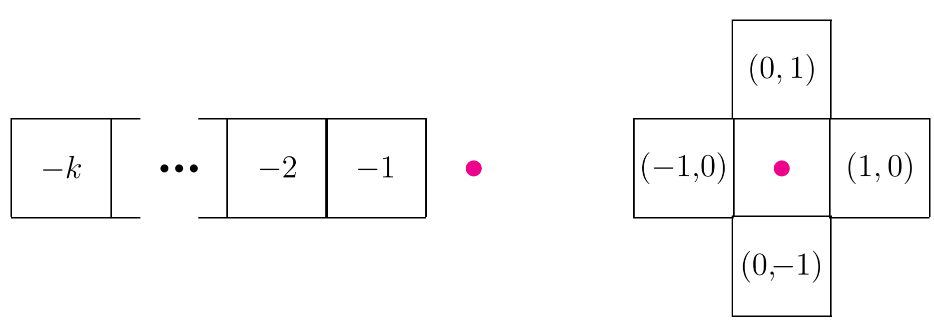
Domain : An index set

- Principal Examples:
  Any finite set
  Z_n, the first n integers (e.g., time series)
  Z/Z_n, the cyclic version of Z_n
  R, the real numbers

Domains supporting translation play a special role

- Operators on domains give means of combination
  $\cup$, the union
  $\times$, the cross-product
  Allows formation of domains for images, etc.

- *Stencil* is a Domain identifying a local neighborhood
  $\{-k, ..., -2, -1\}$, for a k-order autoregressive model
  $\{(-1, 0), (1, 0), (0, -1), (0, 1)\}$, for a first-order MRF



SPATIO-TEMPORAL MODELING (II)

Field : Mapping on a Domain

Random Field a mapping from a Domain to earlier Variables
...the spatiotemporal generalization of random variable

Principal examples:
Time series are random fields over Z or R
Multispectral images: random fields over $\times(\{1, \ldots k\}, Z_n, Z_n)$
(spectral index does not support translation)

Neighborhood a Field from (Domain, Stencil) to a Domain
...maps $(\text{site}, \text{offset}) \mapsto \text{site}'$, often by translation
...supports *adjacency* for dependence structures

Let M be the neighborhood corresponding to the order-1 MRF
Then $M(i, k)$ is the k-th neighbor of site i
$M(i)$ is the set of all neighbors of site i

unpack operator
...returns the neighborhood M given a Domain and Stencil

MODEL SPECIFICATION

- Simplest models have no conditional dependence:

$$\mathcal{D} = Z_n$$
$$(\forall i \in \mathcal{D})\, x[i] \sim \text{Normal}(i, 4)$$

- AR model:

$$\mathcal{D} = Z_n$$
$$\mathcal{S} = -1$$
$$M = \text{unpack}(\mathcal{D}, \mathcal{S})$$
$$(\forall i \in \mathcal{D})\, x[i] \sim \text{Normal}(x[M(i; -1)], 1)$$

- The standard Potts MRF prior:

$$\mathcal{D} = \times(Z_n, Z_n)$$
$$\mathcal{S} = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$$
$$M = \text{unpack}(\mathcal{D}, \mathcal{S})$$
$$(\forall i \in \mathcal{D})\, ct[i] = \sum_{k \in M(i)} x[M(i; k)]$$
$$(\forall i \in \mathcal{D})\, x[i] \sim \text{Discrete}(0, \frac{e^{ct[i]-4}}{e^{ct[i]-4} + e^{-ct[i]}}, 1, \frac{e^{-ct[i]}}{e^{ct[i]-4} + e^{-ct[i]}})$$

Import just enough mathematical notation to express the models

CONCLUSIONS

Machine procedures offer many benefits to scientific inference

Persistent issues:
  Building tractable models of observational reality
  Obtaining accurate training data
  Designing and executing clear falsification experiments

Finding Objects

Discussed a good algorithm-based approach
Divide and conquer schema applicable (even suggestive) here

Labeling Images

Use of statistical models allows falsification experiments,
  easy extension to wider class of problems
Spatially, temporally uniform data is key to accurate labelings

Complex models

Useable temporal and spatial statistical models do exist
...but the best ML perspectives often absent from this work
  agnostic models, robust algorithms, cross-validation, automation
Cooperating space/time models, linked spatiotemporal models

Futures

Object-level recognition in non-algorithmic framework
Languages to express statistical models on structured domains
Model selection in complex, flexible model space